



## Fusing Visual and Clinical Information for Lung Tissue Classification in HRCT Data

Adrien Depeursinge, Daniel Racocceanu, Jimison Iavindrasana, Gilles Cohen, Alexandra Platon, Pierre-Alexandre Poletti, Henning Muller

### ► To cite this version:

Adrien Depeursinge, Daniel Racocceanu, Jimison Iavindrasana, Gilles Cohen, Alexandra Platon, et al.. Fusing Visual and Clinical Information for Lung Tissue Classification in HRCT Data. Artificial Intelligence in Medicine, 2010, pp.ARTMED1118. hal-00493108

**HAL Id: hal-00493108**

**<https://hal.science/hal-00493108>**

Submitted on 18 Jun 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Fusing Visual and Clinical Information for Lung Tissue Classification in HRCT Data

Adrien Depeursinge <sup>a</sup>, Daniel Racocceanu <sup>b</sup>, Jimison Iavindrasana <sup>a</sup>,  
Gilles Cohen <sup>a</sup>, Alexandra Platon <sup>c</sup>, Pierre–Alexandre Poletti <sup>c</sup>,  
Henning Müller <sup>a,d</sup>

<sup>a</sup>Medical Informatics Service, Geneva University Hospitals and University of Geneva (HUG), Geneva, Switzerland

<sup>b</sup>IPAL International Joint Lab, Institute for Infocomm Research (I2R), Centre National de la Recherche Scientifique (CNRS), Singapore

<sup>c</sup>Emergency Radiology Service, HUG, Geneva, Switzerland

<sup>d</sup>University of Applied Sciences Western Switzerland (HES SO), Sierre, Switzerland

## Abstract

*In this paper, we investigate the influence of the clinical context of high–resolution computed tomography (HRCT) images of the chest on tissue classification. 2D regions of interest (ROI) in HRCT axial slices from patients affected with an interstitial lung disease (ILD) are automatically classified into five classes of lung tissue. Relevance of the clinical parameters is studied before fusing them with visual attributes. Two multimedia fusion techniques are compared: early versus late fusion. Early fusion concatenates features in one single vector, yielding a true multimedia feature space. Late fusion consisting of the combination of the probability outputs of two support vector machines (SVM) allowed a maximum of 84% correct predictions of testing instances among the five classes of lung tissue. This represents a significant improvement of 10% compared to a pure visual–based classification. Moreover, the late fusion scheme showed high robustness to the number of clinical parameters used, which suggests that it is appropriate for mining clinical attributes with missing values in clinical routine.*

**Key words:** Multimodal information fusion, Computer–aided diagnosis, Contextual image

# 1 Introduction

The interpretation of high-resolution computed tomography (HRCT) images of the chest from patients affected with interstitial lung diseases (ILDs) is challenging and time-consuming even for experienced radiologists. The term interstitial lung disease accounts for around 150 illnesses of which many forms are rare. Images play an important role for confirming the diagnosis and patients may not require surgical lung biopsy when the clinical and radiographic (HRCT) impression is consistent with a safe diagnosis [15]. The most common imaging procedure used is the chest x-ray because of its low cost and weak radiation exposure. However, chest x-rays are negative in a large proportion of diseases and often unspecific where HRCT of the chest contains essential visual data for the characterization of lung tissue patterns associated with ILDs [36]. The three-dimensional form of HRCT data requires significant reading time, effort, and experience for a correct interpretation. Owing to this intrinsic complexity of the interpretation of HRCTs, an image-based computerized diagnostic aid tool (CAD) can bring quick and precious information, particularly to less experienced radiologists and non-chest experts [2, 13, 33]. Moreover, a radiologist's ability to interpret HRCT images is likely to change based on the domain-specific experience, human factors and time of the day where computerized image analysis is 100% reproducible.

## 1.1 The clinical context of HRCT images

When analyzing an image, one interprets its content according to a given context. This is particularly true when analyzing medical images. Radiologists do never interpret HRCT images without taking into account the clinical context. For example, discovering some fibrotic findings in a lung belonging to an 80-year-old patient is not as surprising as finding some in a lung of a 25-years-old young person. Several clinical parameters — in particular the age of the patient (see Fig. 1) — have a major influence on the visual aspect (density) of HRCT images of the chest [26, 32]. In Figure 1, one can see that healthy tissue from the 88-year-old man has lower mean density with more pre-fibrotic lesions compared to the homogeneous healthy tissue of the 25-year-old man. To accurately analyze HRCT images an image-based computerized diagnostic aid system for ILDs must integrate the clinical context of the images.

## 1.2 Contextual image analysis

Although fundamental in almost every medical field, the context is rarely used in computer vision applications. On the one hand, collecting contextual information beside images is usually time-consuming, requiring the help of a specialist. On the other hand, a high-level of knowledge of the application domain is required to find relevant contextual parameters [37]. The selection of parameters for contextual

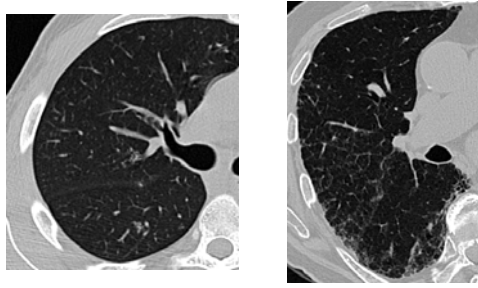


Fig. 1. Healthy tissue from a 25-year-old man on the left, and from an 88-year-old man on the right. Both images have identical window level settings.

medical image analysis has to be carried out based on domain-specific literature along with knowledge bases of computer-based diagnostic decision support systems.

Contextual image analysis implies the fusion of multimodal information sources. When carrying out contextual image analysis of lung tissue in HRCT data, the visual information extracted from HRCT image series is combined with the clinical parameters of the corresponding patient. Integrating information from multiple modalities consists of two major steps [43]. Firstly, the best modalities have to be identified. The best modalities have to be informative according to the considered classes along with being complementary among each other. Each modality  $k$  is represented by a set of features  $\mathbf{v}_k$ . Secondly, the information from the best modalities must be combined with an optimal scheme in order to allow for synergy. The so-called "fusion" can be carried out according to two main strategies [16, 35]:

- early fusion, where features are concatenated into one vector  $\mathbf{v} = (\mathbf{v}_1 \dots \mathbf{v}_k)$  to create one unique feature space
- late fusion, where multiple classifiers  $h_k$  are built on each modality  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  and the fusion is carried out at the decision level

Early fusion allows for a true multimedia (images and clinical data) representation. One single classifier can learn from all information sources. However this method is confronted to the *curse of dimensionality* because the dimension of the resulting feature space  $\mathbf{v}$  is equal to sum of the dimensions of the subspaces  $\mathbf{v}_k$ . Even associated with feature weighting, high-dimensional spaces tend to scatter the homogeneous clusters of instances belonging to the same class. This is particularly true when negative synergies occur among features [3, 22].

Combining classifiers has been a very active domain over the past ten years [21, 24]. Wide interest for the latter mainly relies on the assumption that the heterogeneity of classifiers  $h_k$  leads to better results [12]. Late fusion is a special case of classifier combination where heterogeneity occurs in the input spaces. It allows for a reduction of the data dimensionality by dividing the feature space without neglecting information contained in features that would be discarded by classical feature selection methods [34].

### 1.3 Related work

Context has been used in content-based image retrieval (CBIR) where information from textual annotations of images was fused with image features [41] (i.e. grey-level histograms and texture features). In [5], a CBIR system combined visual statistics with textual statistics directly in the feature vector space representation. Inter-media medical image retrieval was carried out in [23] using textual features semantically parsed and described with the Unified Medical Language System (UMLS) along with color and texture features. The visual and textual information was combined in the calculation of the similarity measure. Investigation of the effectiveness of combining text and image for retrieval including medical image retrieval is one of the main goals of the CBIR benchmarking campaign ImageCLEF<sup>1</sup> [17, 27].

Combined decision of classifiers constructed on sequentially selected sets of features were tested on four datasets including medical data in [34]. Best results were obtained when the combination approach was applied on top of feature selection. Unfortunately, experiments were carried out with homogeneous datasets which did not contain heterogeneous features such as visual, textual, audio, etc.

Early fusion of clinical parameters and genetic factors was used to predict the risk of coronary artery disease in [7]. Interaction among features were studied using Bayesian network representations. Although visually identified feature groups were in accordance with medical knowledge, no quantitative analysis of the interactions was performed.

A combination of radiologic findings on chest radiographs and clinical parameters to provide probability output of 11 possible ILDs using an artificial neural network is carried out in [1]. By using these probabilities, radiologists were able to significantly improve their diagnosing accuracies. However, automatic detection of relevant patterns in the chest radiographs was not investigated.

Utilization of knowledge of disease location to improve detection of *fibrosis* patterns in HRCT data was carried out in [44]. The locations of the patterns showed to significantly improve detection performance but require an accurate segmentation of the anatomy of the lung.

Many image-based diagnostic aid systems for ILDs achieved high classification accuracy of lung tissue patterns in HRCT data [33, 38, 39] and showed to be effective in clinical routine [2]. Yet, most of these systems are based on visual data, only (HRCT images). To our knowledge no system attempted to integrate clinical parameters for automatic detection of lung tissue patterns associated with ILDs in HRCT data. Texture analysis of lung images using Wavelet frames was investigated in [9, 11] and support vector machines (SVM) showed to be optimal for the categorization of lung tissue using quincunx Wavelet frames in [10]. In this paper, we study the influence of the integration of the clinical context of HRCT images on classification performance of 2D regions of interest (ROI) drawn by two radiologists in axial slices from patients affected with an ILD.

The paper is structured as follows. In Section 2, the dataset and software used for evaluating the influence of the clinical context on lung tissue classification in HRCT images is described. Section 3 is divided into two parts. Section 3.1 describes the composition of the multimodal feature space whereas

---

<sup>1</sup> <http://www.imageclef.org/>

the comparison of the fusing techniques is carried out Section 3.2. Results are interpreted and discussed in Section 4, future work is proposed in Section 5 and final conclusions are drawn in section 6.

## 2 Methods

The dataset used to investigate the influence of clinical parameters on classification accuracy of lung tissue is part of an internal multimedia database of ILD cases containing HRCT images created in Talisman<sup>2</sup>. 99 relevant clinical parameters were chosen according to the 15 most frequent ILDs [18] based on the literature [20, 40], along with knowledge bases of computer-based diagnostic decision support systems. Discussions and remarks from lung specialists, radiologists and the medical informatics service (SIM<sup>3</sup>) at the University Hospitals of Geneva (HUG) allowed an iterative review of the selected parameters as well as standardized units and data format to be used. The parameters that were not available from the electronic health record (EHR) were removed. An HTML form and PHP scripts were used to collect the clinical parameters and to store them into a MySQL database (see Fig. 2). When multiples instances of clinical parameters (e.g. laboratory data) were available in the EHR, the instance as close as possible to HRCT examinations was retained. 96 patients with confirmed diagnosis were retrospectively collected at the HUG between 2003 and 2006. For each patient, a physician filled as many clinical parameters as possible and a total of 1104 ROIs of lung tissue patterns were drawn in full-resolution DICOM images by two experienced radiologists. The slice thickness of the images is limited to 1mm. A graphical user interface implemented in Java was developed in order to meet the needs of the radiologists for the various annotation tasks.

736 ROIs from healthy and four pathologic lung tissue patterns belonging to 48 patients with filled clinical parameters were selected for this study (see Table 1). Patterns that are represented by less than 4 patients are left aside. The selected patterns are *healthy*, *emphysema*, *ground glass*, *fibrosis* and *micronodules*. Distributions of the classes are highly imbalanced as the largest class *fibrosis* contains 312 ROIs and the smallest class only 58 ROIs. There is a mean of 147.2 ROIs per class.

Implementation of the SVMs'  $C$ -support vector classification is taken from the open source Java library *Weka*<sup>4</sup> using a wrapper for *LIBSVM*<sup>5</sup>. The image feature extraction and the optimization of SVMs is implemented in Java. Quincunx Wavelet frames are implemented in Java [14].

## 3 Results

The first part of this section describes the composition of the features in each modality used for the classification of the lung tissue patterns. The second part describes and details the results of the fusion

<sup>2</sup> TALISMAN: Texture Analysis of Lung ImageS for Medical diagnostic AssistaNce, [http://www.sim.hcuge.ch/medgift/01\\_Talisman\\_EN.htm](http://www.sim.hcuge.ch/medgift/01_Talisman_EN.htm)

<sup>3</sup> <http://www.sim.hcuge.ch/>

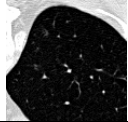
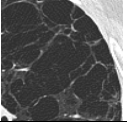
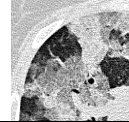
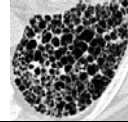

<sup>4</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

<sup>5</sup> <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

Fig. 2. A screenshot of the form used for data entry of the 99 clinical parameters.

Table 1

Distribution of the ROIs and patients per class of lung tissue pattern. Each patient may contain several kinds of lung tissue patterns.

visual aspect					
class	healthy	emphysema	ground glass	fibrosis	micronodules
# of ROIs	63	58	148	312	155
# of patients	5	4	14	28	5

of the two modalities.

### 3.1 Modalities

The two modalities used for classifying the lung tissue patterns are made up of the visual information from HRCT images of the chest as well as the corresponding textual information describing the clinical state of the patient at the time of the disease episode.

### 3.1.1 Clinical features $c_n$

The clinical parameters entered in the MySQL database are not directly usable for data-mining. Pre-processing steps are required to build a workable feature space. Nominal variables are divided into binary features. Textual variables and binary variables that contained one single modality are left aside. Since leaving aside cases with missing values is not conceivable, variables with less than 50% of the values filled were removed and average values were substituted. Note that a filled value means not an *unknown* value which is used when the clinical parameter was not detailed in the EHR. After having gathered binary and continuous variables, the created clinical feature space contains 72 attributes (63 binary and 9 continuous). For example, over the 48 selected patients, the parameter *host\_HIV* has 3 *yes*, 43 *no* and 2 *unknown* values. *yes* values are coded with 1, *no* with 0 and thus the missing values are substituted by the mean: 0.065. The continuous features were not discretized as it is preferable not to group their values together into categories with a further purpose of separating clusters of instances in the feature space. The mean filling rate of the retained attributes is 88.7%.

### 3.1.2 Visual features $t_m$

Visual features consist of texture features with 22 bins of grey-level histograms of Hounsfield Units (H.U.) within the ROIs, along with quincunx Wavelet frame (QWF) coefficients extracted at 8 scales. The distributions of the wavelet coefficients in each subband  $i$  are characterized through the parameters of mixtures of two Gaussians. With fixed means  $\mu_{1,2}^i = \mu^i$ , the standard-deviations  $\sigma_{1,2}^i$  ( $\sigma_1^i > \sigma_2^i$ ) are estimated using the expectation-maximization (EM) algorithm. An additional feature *airpix* measuring the number of pixels of the ROI with value inferior to  $-1000$  H.U. (which corresponds to the density of air) is used. In total, the visual feature set is composed of 47 attributes. A complete description and evaluation of the visual feature space can be found in [9].

## 3.2 Combining features: early versus late fusion

In order to study the effect of the integration of the clinical context of HRCT images on the classification accuracy of the lung tissue patterns, optimized SVMs with a Gaussian kernel are used to categorize ROIs from the multimodal feature space. SVMs with a Gaussian kernel have shown to be effective to categorize lung tissue patterns from visual features in [10] and are adapted to mine clinical parameters as shown in [8]. Two methods for combining visual and clinical attributes are compared: early versus late fusion (Sections 3.2.2 and 3.2.3). The relevance of clinical attributes is studied in Section 3.2.1.

### 3.2.1 Ranking the clinical attributes

Integrating the clinical context in lung tissue classification implicitly assumes that clinical parameters contain relevant information to predict the types of lung tissue contained in HRCT image series of a patient affected with an ILD. Although parameters such as *age* are clearly related to the visual aspect of the lung tissue (see Section 1.1), dependencies between clinical attributes and classes of lung tissue must be investigated before any fusion with the visual features. Indeed, due to missing values, binarization or



irrelevance according to the studied diseases, some features might introduce noise by scattering homogeneous clusters of instances in the feature space. High presence of binary attributes increases the risk of obtaining XOR configurations of instances, which leads to highly non-linear decision boundaries. Moreover, it is preferable to keep as few as possible features to limit the curse of dimensionality, especially for early fusion. A feature ranking is thus required to build an effective set of attributes.

Two measures are compared in their ability to rank the clinical attributes for lung tissue classification: the information gain ratio  $I_{G_{ratio}}$  and the single testing accuracy  $A^{single}$ .  $I_{G_{ratio}}$  is derived from the information gain measure  $I_G$  originally used by Quinlan in decision trees in [30]. The information gain  $I_G(Y|X)$  of a given attribute  $X$  with respect to the class attribute  $Y$  quantifies the change in information entropy when the value of  $X$  is revealed:

$$I_G(Y|X) = H(Y) - H(Y|X) \quad (1)$$

The information entropy  $H(Y)$  measures the uncertainty about the value of  $Y$  and the conditional information entropy  $H(Y|X)$  measures the uncertainty about the value of  $Y$  when the value of  $X$  is known:

$$H(Y) = - \sum_{y \in \mathcal{Y}} p(y) \log p(y) \quad (2)$$

$$H(Y|X) = - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (3)$$

The information gain ratio  $I_{G_{ratio}}$  is derived from  $I_G$  using

$$I_{G_{ratio}}(Y|X) = \frac{I_G(Y|X)}{- \sum_{i=1}^l \frac{|T_i|}{|T|} \log(\frac{|T_i|}{|T|})} \quad (4)$$

with  $T$  the training set and  $l$  the number of possible values of  $X$ . Compared to  $I_G$  the gain ratio will not give advantage to attributes with a high range of possible values [29,31]. As the clinical feature space is populated with binary as well as continuous attributes, it is highly preferable to use the  $I_{G_{ratio}}$  for ranking. Another measure proposed for ranking the clinical attributes is the single testing accuracy  $A^{single}$ .  $A^{single}$  is defined as the classification accuracy with SVMs (see Section 3.2.4) based on a feature vector concatenating all visual features  $t_m$  with the studied clinical feature  $c_n$ :

$$\mathbf{v} = (t_1 \dots t_M c_n) \quad (5)$$

The correlation of  $I_{G_{ratio}}$  and  $A^{single}$  obtained using the experimental setup described in Section 3.2.4 is studied in Figure 3. Table 2 lists the first 20 clinical attributes with highest  $A^{single}$  value. The correlation matrix of the feature space containing the visual features along with the first 20 clinical attributes is shown in Figure 4.

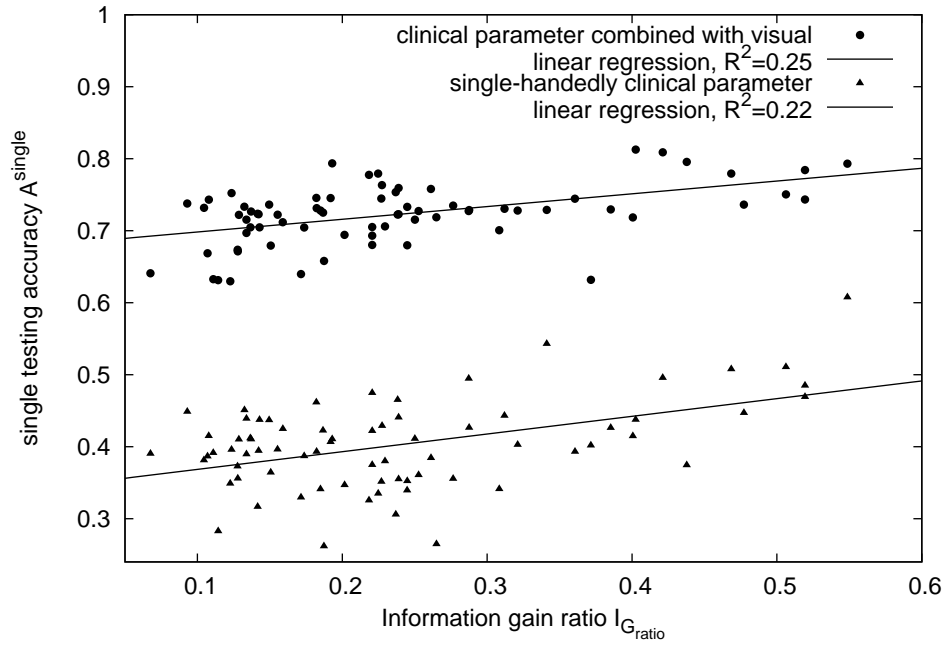


Fig. 3. Correlation of  $I_{G_{ratio}}$  and  $A^{single}$ .

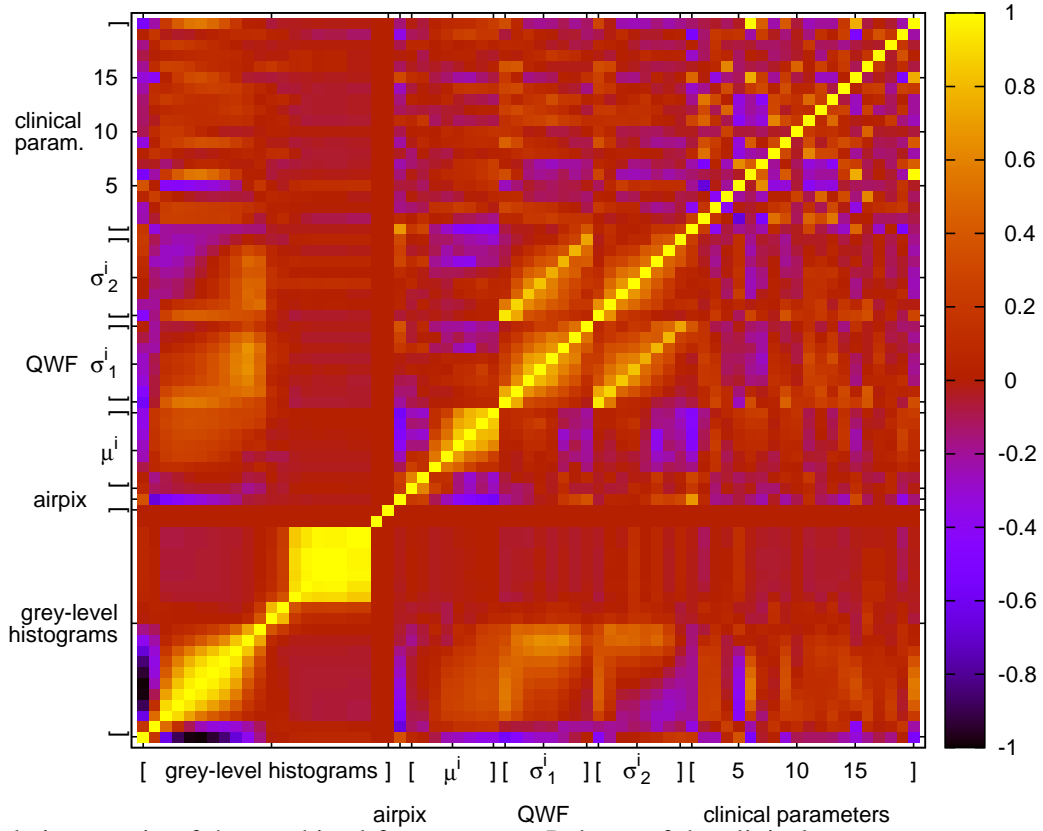


Fig. 4. Correlation matrix of the combined feature space. Indexes of the clinical parameters corresponds to their rank described in Table 2.

Table 2

List of the first 20 clinical attributes with highest  $A^{single}$  when combined with visual features. Abbreviations: HTA: arterial hypertension, subOAP: acute pulmonary edema, LDH: serum lactate dehydrogenase.

rank	$A^{single}$	$I_{G_{ratio}}$	name	type
1	0.813	0.403	laboratory_hematocrit	continuous
2	0.809	0.421	age	continuous
3	0.796	0.438	laboratory_hemoglobin	continuous
4	0.794	0.193	past_medical_allergy	binary
5	0.793	0.549	findings_physical_generals_lymph	binary
6	0.784	0.519	past_medical_lymphom	binary
7	0.779	0.469	findings_physical_generals_fever	binary
8	0.779	0.225	medication_cordarone	binary
9	0.778	0.218	host_diabetes	binary
10	0.763	0.227	biopsy_bronchoscopy_transbronchial_eosinophil	binary
11	0.759	0.239	past_medical_HTA	binary
12	0.758	0.261	past_medical_dyspnea_attack	binary
13	0.754	0.237	past_medical_subOAP	binary
14	0.752	0.124	findings_physical_respiratory_tachypnea	binary
15	0.751	0.506	host_chemotherapy	binary
16	0.746	0.182	past_medical_wheezing	binary
17	0.745	0.192	findings_physical_abdominals_liver	binary
18	0.745	0.227	biopsy_bronchoscopy_transbronchial_interstitial_fibrosis	binary
19	0.745	0.361	laboratory_LDH	continuous
20	0.743	0.52	host_hemopathy	binary

### 3.2.2 Early fusion: feature concatenation

In order to create a multimodal feature space, clinical attributes and visual features are normalized and concatenated into one single feature vector  $\mathbf{v}$  as follows:

$$\mathbf{v} = (t_1 \dots t_M \ c_1 \dots c_N) \quad (6)$$

with  $t_m, m \in [1; M]$  the visual features and  $c_n, n \in [1; N]$  the clinical attributes. Using all 72 clinical attributes, the maximum dimensionality of the multimodal feature space reaches 119 with 56 continuous and 63 binary features.  $\mathbf{v}$  is used as input of SVMs which directly output the predicted class using one versus one multiclass approach. Testing accuracies are obtained using the experimental setup described

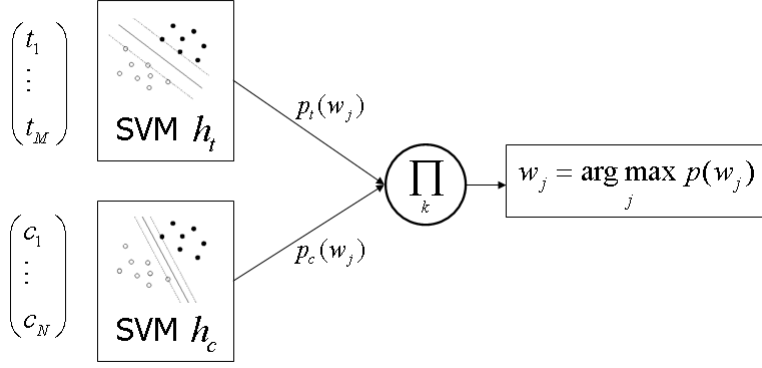


Fig. 5. Classification scheme for late fusion. Two expert SVM classifiers  $h_t$  and  $h_c$  output probabilities  $p_k(w_j)$  which are multiplied to obtain the final probability of each class  $w_j$ .

in Section 3.2.4.

### 3.2.3 Late fusion: Combination of SVM classifiers

Two SVM classifiers  $h_t$  and  $h_c$  are trained using visual features  $\mathbf{t}$  and clinical attributes  $\mathbf{c}$  respectively. Attributes in  $\mathbf{t}$  and  $\mathbf{c}$  are normalized (within each group) in order to give equal importance to each of them. Both SVM output probabilities  $p_t$  and  $p_c$  using pairwise coupling [42]. For each class  $w_j$  ( $j = 1 \dots 5$ ), probabilities are multiplied to compute final probability  $p(w_j) = p_t(w_j) \cdot p_c(w_j)$ . The final predicted class  $w_j$  is given by  $\arg \max_j p(w_j)$ . Using the product of probabilities for predicting the final class assumes that the modalities  $\mathbf{t}$  and  $\mathbf{c}$  are conditionally statistically independent [4, 21], which is admissible as mean correlation value  $\rho_{mean}$  of each feature pair  $(t_m, c_n)$  is equal to 0.0143 (see Figure 4). Late fusion scheme is summarized in Figure 5.

Smaller feature subspaces have the advantage to reduce the computational complexity for solving the quadratic problem of finding the maximum margin hyperplane of SVMs. Moreover the subspaces can be processed in parallel to allow for faster training of the SVMs.

### 3.2.4 Experimental setup

In order to test the influence of clinical parameters on classification accuracy of the 2D ROIs, a leave-one-patient-out cross-validation was used. The latter is in accordance with a clinical usage of the CAD. Each of the ROIs belonging to one same patient are left aside for testing and all remaining ROIs are used to train and optimize the SVMs. Indeed, the training set is used both for grid search for optimal parameters and adjustment of the maximum-margin hyperplane of the SVMs. Optimized parameters of the SVM are the cost of the errors  $C$  and the width  $\sigma$  of the Gaussian kernel. A grid search is carried out within the intervals  $C \in [1; 100]$  and  $\sigma \in [10^{-2}; 10^2]$ . For every coordinate of the grid, a 10-fold cross-validation (CV) is carried out on the training set. Optimal parameters  $(C_{opt}, \sigma_{opt})$  that allowed best mean CV accuracy  $A^{cv}$  are used to train the final model on the entire training set. A preliminary coarse

Table 3

Averaged accuracies obtained with the various techniques. Best performances are highlighted in bold.

	visual features	clinical features	concatenated features	combined SVMs	combined SVMs, PCA on clinical features	PCA on concatenated features
healthy	0.46	0.01	0.19	0.43	<b>0.48</b>	0.22
emphysema	0.64	0.08	0.26	0.43	0.47	<b>0.78</b>
ground glass	0.57	0.56	0.6	<b>0.71</b>	0.62	0.5
fibrosis	0.91	0.77	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>	0.87
micronodules	0.68	0.8	<b>0.86</b>	0.83	0.45	0.21
global	0.74	0.61	0.74	<b>0.79</b>	0.72	0.58

grid search was performed to locate regions of the space with high  $A^{cv}$  values.

As the toughness of the classification task can strongly vary depending on the drawdown of the testing patient, the global experimentation is repeated for all 48 patients to obtain reliable accuracy values. In order to study the optimal number  $n$  of clinical attributes to be used, mean classification accuracies over the 48 patients are computed for each  $n \in [1; 72]$ , with clinical attributes ordered by  $A^{single}$  values (see Section 4.1). Mean classification accuracies according to  $n$  obtained with the test set using visual features only, clinical features only and combined features with early and late fusion are shown in Figures 6, 7, 8, 9, 10 for each class. Global accuracies of each method are summarized in Figure 11. Mean accuracies over  $n$  values as well as classification based on principal component analysis (PCA) are contained in Table 3. First, a PCA transform is applied only on the clinical feature set  $c_n$  to be used with combined SVMs. Second, a PCA transform is applied on the whole concatenated feature set. The number of principal components  $P$  kept is chosen according to [28]:

$$P > 1 + 2\sqrt{\frac{N-1}{K-1}} \quad (7)$$

with  $N$  the number of features and  $K$  the number of instances. In both cases,  $P = 2$  is chosen based on (7).

## 4 Interpretation

The first part of this section discusses the limitations of the ranking measures  $I_{Gratio}$  and  $A^{single}$ . The second part verifies the relevance of the 20 clinical attributes with highest  $A^{single}$  value to the medical domain. The third part studies the consistency and the complementarity of the multimodal feature space through the correlation matrix presented in Figure 4. The last section interprets the performances of classification of the two fusion designs presented in Table 3 and Figures 6-11.

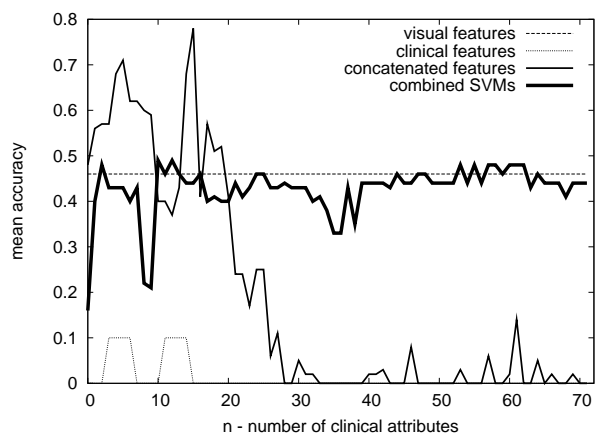


Fig. 6. Classification accuracies for class *healthy*.

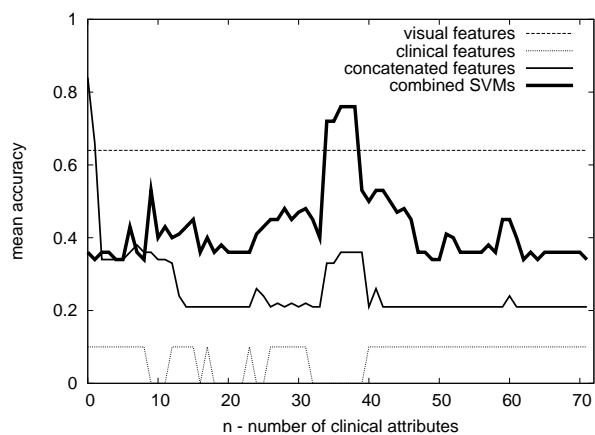


Fig. 7. Accuracies for class *emphysema*.

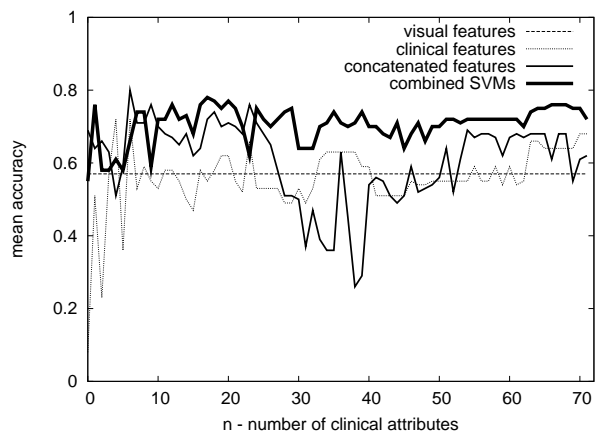


Fig. 8. Accuracies for class *ground glass*.

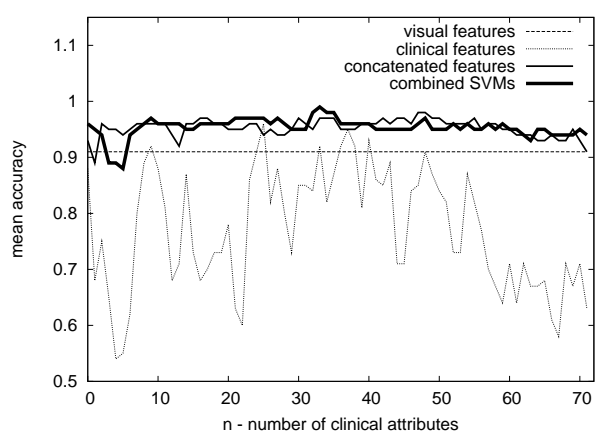


Fig. 9. Accuracies for class *fibrosis*.

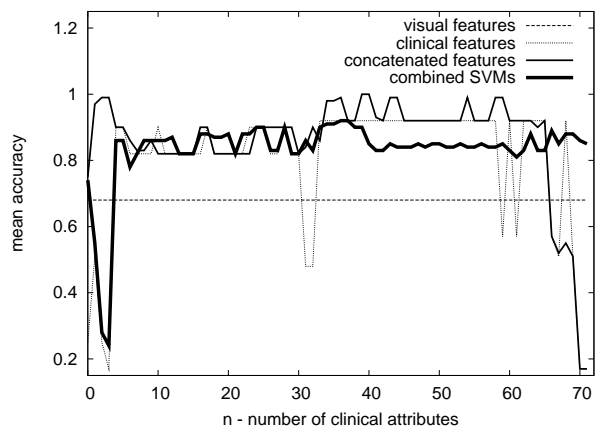


Fig. 10. Accuracies for class *micronodules*.

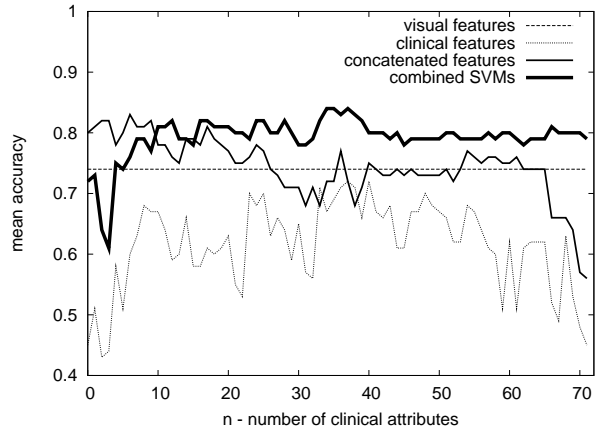


Fig. 11. Global classification accuracies.

## 4.1 Measures for ranking

Figure 3 shows that  $I_{G_{ratio}}$  is little correlated to both  $A^{single}$  and testing accuracy obtained with each single-handedly clinical parameter. Pearson’s coefficient of regression  $R^2$  is below 0.25 for both comparisons. Even if  $A^{single}$  is averaged over 30 experiments, the values obtained still have high variance according to the drawdowns of the training and testing sets. This is a first explication for having low values of  $R^2$ . A second explication comes with the definition of  $I_{G_{ratio}}$  which measures the relevance of each separated single attribute. One feature generally distinguishes classes in combination with other features [19], which suggests that  $I_{G_{ratio}}$  is not convenient to rank the clinical attributes with a purpose of fusing them with visual features.  $I_{G_{ratio}}$  is also known to be unstable as it is very sensitive to small changes in the training set [12] which is not desirable for ranking attributes from a high-dimensional set of heterogeneous features. Due to the several drawbacks of  $I_{G_{ratio}}$ ,  $A^{single}$  is used for ranking the clinical attributes.

## 4.2 Relevance of clinical attributes

The relevance of the clinical attributes for classifying lung tissue patterns in HRCT data is subject to many external factors such as the availability of the parameters in the EHR, its binarization required to be added to  $\mathbf{v}$  and relevance according to the studied diseases. Indeed a parameter such as the result of a lung biopsy is obviously highly informative for characterizing the lung tissue but is rarely carried out and available in the EHR. The categorization and binarization has also major influence on the quality of clinical data. At last, the relevance of the parameter according to the studied diseases is of course primordial.

As observed in Section 1.1, the age has an important influence on the visual aspect of lung tissue (see Figure 1) and this is confirmed by finding it at the 2<sup>nd</sup> rank in Table 2. The presence of the parameter *laboratory\_hematocrit* at the top of the list is a bit more subtle. An explication for this is that large homogeneous regions of air, characterizing *emphysema* patterns, will cause hypoxia and may elicit an increased production of red blood cells by the kidney, and thus increase the level of hematocrit. This phenomenon is indeed commonly observed in cases affected with chronic obstructive pulmonary disease (COPD) [6], characterized by HRCT images showing *emphysema* patterns. The latter observation is firmly confirmed by looking at the correlation matrix in Figure 4, where the first clinical parameter (*laboratory\_hematocrit*) is strongly anticorrelated with the means  $\mu_i$  of the QWF and highly correlated to *airpix* the number of pixels of air within the ROIs. Indeed the means  $\mu_i$  of the QWF have high values for inhomogeneous patterns, where *emphysema* patterns are very homogeneous due to absence of lung tissue. Coherently, the 3<sup>rd</sup> rank is occupied by the parameter *laboratory\_hemoglobin* which is also involved in the transport of oxygen. Indeed, hemoglobin is the protein contained in red blood cells that is responsible for delivery of oxygen to the tissues. To ensure adequate tissue oxygenation, a sufficient hemoglobin level must be maintained.

The presences of parameters *findings\_physical\_general\_lymph* (enlargement of lymph node(s)) and *findings\_physical\_general\_fever* at the 5<sup>th</sup> and 7<sup>th</sup> ranks are not surprising as they usually highlights the

presence of a host illness. Finding the parameter *medication\_cordarone* at the 8<sup>th</sup> rank is in accordance with the well-known side effect of the cordarone drug creating pulmonary fibrosis on the long range.

### 4.3 Consistency of the multimodal feature space

The study of the correlation of the multimodal feature space is carried out in Figure 4. A first look at the correlation matrix shows that clinical features have small correlation with the visual features. This is confirmed as mean correlation  $\rho_{mean}$  is equal to 0.0143. Several homogeneous groups can be identified within the visual features. The first histogram bin representing pixels values within  $[-1050; -975[$  H.U. is of course highly correlated with *airpix* and is anticorrelated with bins in range  $[-900; -450[$ . This partly due to patterns with low-density tissues (mostly *emphysema*) are mainly composed by air and thus do contain few pulmonary tissue in range  $[-900; -450[$ . Globally, histograms are logically correlated in contiguous pairs. Bins 14–20 form a strongly correlated group ( $\rho_{mean} = 0.95$ ) which shows that high-density tissues with H.U. values in  $[-75; 375[$  only occur together, most probably in *fibrosis* and *ground glass* patterns. Bins 21 and 22 are very sparse and thus not correlated to any other attribute. Within the QWF features, two groups can be identified: the means  $\mu^i$  and the standard-deviations  $\sigma_{1,2}^i$ .  $\rho_{mean}$  is equal to -0.04 between the two groups. Within the groups, it is not surprising to observe that means and standard-deviations of two consecutive Wavelet subbands are correlated.

Within the clinical parameters, 3 are highly correlated: *past\_medical\_lymphom*, *host\_chemotherapy* and *host\_hemopathy* ( $\rho_{mean} = 0.84$ ). This is not surprising as *past\_medical\_lymphom*, which stands for having had a lymphoma or leukemia, which is a type of hemopathy (blood cancer), is treated with chemotherapy. Those three parameter are all involved in ILDs as chemotherapy can induce diverse injuries of the lung tissue [25].

### 4.4 Influence of the clinical context on lung tissue classification: early versus late fusion

Influence of the clinical context of HRCT images on lung tissue classification accuracy is studied in Figures 6, 7, 8, 9, 10 and 11. As baseline performance, let us consider the accuracy achieved by using visual features only which has a mean global value of 74%. The last line of Table 3 shows that integrating the clinical context of the images allows for significant global improvements of the classification accuracies. In the mean, 5% is gained in global accuracy using combined SVMs compared to using visual features only. Classification accuracies of combined modalities are in mean always superior to single modalities as observed in Table 3. However, the clinical features can harm the classification accuracy if they are not integrated using an appropriate fusing technique. With early fusion, positive interactions between clinical and visual features are allowed as showed in Fig. 10 and for low  $n$  values in Fig. 6. However, concatenating all features in a single vector has the drawback that less informative attributes scatter homogeneous clusters of instances in the feature space. Figures 6, 7 and 11 confirm that phenomenon where the curve of the concatenated features drops when adding more noisy clinical attributes with low discriminatory power. Some clinical features have negative interactions among them as well as with the visual ones. Separating visual and clinical features for mining using late fusion avoid interactions between the feature groups which show more stable performances compared to early fusion. This



is particularly true when clinical attributes carry low information as it is the case for classes *healthy*. Moreover, the combined SVMs show high robustness towards the number  $n$  of clinical parameters used and allowed the best accuracy of 84% correct predictions of testing instances (ROIs) among the five classes of lung tissue with an optimal number of clinical attributes  $n = 35$ . Although PCA transform allows best results for classes where clinical attributes have low discriminatory power (see Table 3), it does not improve the global accuracy.

## 5 Future work

We believe that negative synergies still occur among features using a late fusion scheme. Groups of features with positive synergy [3] which allows for homogeneous cluster of instances belonging to the same class must be identified and mined into separated subspaces. Indeed, part of the fluctuations of the performances according to  $n$  are the result of interactions among the various groups of features. An approach for identifying feature groups with positive synergy based on mutual information is described in [22]. Identifying the groups should also include medical knowledge. A visual approach based on Bayesian networks is proposed in [7].

## 6 Conclusions

In this paper, the influence of the clinical context on lung tissue classification from HRCT data is investigated. Correlation analysis of the multimedia feature space shows that the dataset is in accordance with medical knowledge. Two fusion scheme of the modalities are studied: early versus late fusion. The combination of two SVM classifiers achieved highest classification accuracies and allowed a mean of 79% and a maximum of 84% correct predictions of testing instances among the five classes of lung tissue. This represents a significant benefit of 10% compared to a pure visually-based classification. Late fusion shows robustness towards the number of clinical parameters used, which suggests that it is appropriate for mining clinical attributes with missing values. Accuracy values are trustworthy for further usage in clinical routine as we never train and test with ROIs that belongs to the same patient. In addition, the leave-one-patient-out cross-validation testing approach is in accordance with a clinical usage of the CAD where the radiologist analyzes one patient at a time and the system was trained with all the previously analyzed patients. We believe that the late fusion scheme can still be improved by identifying group of attributes with high synergy and mining them separately in order to preserve homogeneous clusters of instances in the feature space. Moreover, the combination rules among the modalities must be investigated in order to allow optimal complementarity of the modalities.

## 7 Acknowledgments

This work was supported by the Swiss National Science Foundation (FNS) with grant 200020-118638/1, the equalization fund of the University and Hospitals of Geneva (grant 05-9-II), the EU

## References

- [1] H. Abe, K. Ashizawa, F. Li, N. Matsuyama, A. Fukushima, J. Shiraishi, H. MacMahon, and K. Doi. Artificial neural networks (ANNs) for differential diagnosis of interstitial lung disease : results of a simulation test with actual clinical cases. *Academic Radiology*, 11:29–37, 2004.
- [2] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori. Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment. *Radiology*, 228:265–270, 2003.
- [3] A. J. Bell. The co-information lattice. In *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 921–926, Nara, Japan, April 2003.
- [4] R. Benmokhtar and B. Huet. Classifier fusion: Combination methods for semantic indexing in video content. In *ICANN (2)*, pages 65–74, 2006.
- [5] M. L. Cascia, S. Sethi, and S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. In *CBAIVL '98: Proceedings of the IEEE Workshop on Content - Based Access of Image and Video Libraries*, page 24, Washington, DC, USA, 1998. IEEE Computer Society.
- [6] A. Chambellan, E. Chailleux, and T. Similowski. Prognostic value of the hematocrit in patients with severe COPD receiving long-term oxygen therapy. *Chest*, 128:1201–1208, 2005.
- [7] Q. Chen, G. Li, T.-Y. Leong, and C.-K. Heng. Predicting coronary artery disease with medical profile and gene polymorphisms data. In K. A. Kuhn, J. R. Warren, and T.-Y. Leong, editors, *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics*, pages 1219–1224, Brisbane, Australia, August 2007. IOS Press.
- [8] G. Cohen, M. Hilario, H. Sax, S. Hugonnet, and A. Geissbuhler. Learning from imbalanced data in surveillance of nosocomial infection. *Artificial Intelligence in Medicine*, 37(1):7–18, May 2006.
- [9] A. Depeursinge, D. V. de Ville, M. Unser, and H. Müller. Lung tissue analysis using isotropic polyharmonic B-Spline wavelets. In *MICCAI '08: Proceedings of the First International Workshop on Pulmonary Image Analysis*, pages 125–133, New-York, US, September 2008. Springer.
- [10] A. Depeursinge, J. Iavindrasana, A. Hidki, G. Cohen, A. Geissbuhler, A. Platon, P.-A. Poletti, and H. Müller. A classification framework for lung tissue categorization. In K. P. Andriole and K. M. Siddiqui, editors, *Medical Imaging 2008: PACS and Imaging Informatics*, volume 6919, page 69190C. SPIE, April 2008.
- [11] A. Depeursinge, D. Sage, A. Hidki, A. Platon, P.-A. Poletti, M. Unser, and H. Muller. Lung tissue classification using wavelet frames. *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 6259–6262, August 2007.
- [12] T. G. Dietterich. Ensemble methods in machine learning. In *MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems*, pages 1–15, London, UK, 2000. Springer-Verlag.
- [13] K. Doi. Current status and future potential of computer-aided diagnosis in medical imaging. *British Journal of Radiology*, 78:3–19, 2005.
- [14] M. Feilner, D. Van De Ville, and M. Unser. An orthogonal family of quincunx wavelets with continuously adjustable order. *IEEE Transactions on Image Processing*, 14(4):499–510, April 2005.

---

<sup>6</sup> ONCO-MEDIA: ONtology and COntext related MEdical image Distributed Intelligent Access, <http://www.onco-media.com/>

- [15] K. R. Flaherty, T. E. King, J. Ganesh Raghu, J. P. Lynch III, T. V. Colby, W. D. Travis, B. H. Gross, E. A. Kazerooni, G. B. Toews, Q. Long, S. Murray, V. N. Lama, S. E. Gay, and F. J. Martinez. Idiopathic interstitial pneumonia: What is the effect of a multidisciplinary approach to diagnosis? *American Journal of Respiratory and Critical Care Medicine*, 170:904–910, July 2004.
- [16] H. Gunes and M. Piccardi. Affect recognition from face and body: early fusion vs. late fusion. In *2005 IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3437–3443, October 2005.
- [17] W. Hersh, J. Kalpathy-Cramer, and J. Jensen. Medical image retrieval and automated annotation: OHSU at ImageCLEF 2006. In C. Peters, P. Clough, F. C. Gey, J. Karlgren, B. Magnini, D. W. Oard, M. de Rijke, and M. Stempfhuber, editors, *CLEF*, volume 4730 of *Lecture Notes in Computer Science*, pages 660–669. Springer, 2006.
- [18] A. Hidki, H. Müller, A. Depeursinge, P.-A. Poletti, and A. Geissbuhler. Putting the image into perspective: The need for domain knowledge when performing image-based diagnostic aid. In *Swiss conference on medical informatics (SSIM 2006)*, Basel, Switzerland, April 2006.
- [19] S. J. Hong. Use of contextual information for feature ranking and discretization. *IEEE Transactions on Knowledge and Data Engineering*, 9(5):718–730, 1997.
- [20] T. E. King. Approach to the adult with interstitial lung disease. *UpToDate*, August, 2004.
- [21] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(3):226–239, March 1998.
- [22] J. Kludas, E. Bruno, and S. Marchand-Maillet. Information fusion in multimedia information retrieval. In *Proceedings of 5th international Workshop on Adaptive Multimedia Retrieval (AMR)*, Paris, France, July 5-6 2007.
- [23] C. Lacoste, J.-P. Chevallet, J.-H. Lim, D. T. H. Le, W. Xiong, D. Racoceanu, R. Teodorescu, and N. Vuillenmot. Inter-media concept-based medical image indexing and retrieval with UMLS at IPAL. In C. Peters, P. Clough, F. C. Gey, J. Karlgren, B. Magnini, D. W. Oard, M. de Rijke, and M. Stempfhuber, editors, *CLEF*, volume 4730 of *Lecture Notes in Computer Science*, pages 694–701. Springer, 2006.
- [24] L. Lam. Classifier combinations: Implementations and theoretical issues. In *MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems*, pages 77–86, London, UK, 2000. Springer-Verlag.
- [25] A. H. Limper. Chemotherapy-induced lung disease. *Clinics in Chest Medicine*, 25:53–64, 2004.
- [26] F. Mitsunobu, T. Mifune, K. Ashida, Y. Hosaki, H. Tsugeno, M. Okamoto, S. Harada, S. Takata, and Y. Tanizaki. Influence of age and disease severity on high resolution CT lung densitometry in asthma. *Thorax*, 56:851–856, 2001.
- [27] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. *International Journal of Medical Informatics*, 73:1–23, February 2004.
- [28] M. Pechenizkiy, A. Tsymbal, and S. Puuronen. PCA-based feature transformation for classification: Issues in medical diagnostics. In *CBMS '04: Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems*, page 535, Washington, DC, USA, 2004. IEEE Computer Society.
- [29] P. Perner, T. P. Belikova, and N. I. Yashunskaya. Knowledge acquisition by symbolic decision tree induction for interpretation of digital images in radiology. In *SSPR '96: Proceedings of the 6th International Workshop on Advances in Structural and Syntactical Pattern Recognition*, pages 208–219, London, UK, 1996. Springer-Verlag.
- [30] R. J. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, March 1986.
- [31] R. J. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., 1993.
- [32] J. H. Ryu, E. J. Olson, D. E. Midthun, and S. J. Swensen. Diagnostic approach to the patient with diffuse lung disease. In *Mayo Clin Proc.*, volume 77, pages 1221–1227, 2002.

- [33] C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick. ASSERT: A physician-in-the-loop content-based retrieval system for HRCT image databases. *Computer Vision and Image Understanding (special issue on content-based access for image and video libraries)*, 75(1/2):111–132, July/August 1999.
- [34] M. Skurichina and R. P. Duin. Combining feature subsets in feature selection. In *Multiple Classifier Systems*, volume 3541 of *Lecture Notes in Computer Science*, pages 165–175. Springer, 2005.
- [35] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders. Early versus late fusion in semantic video analysis. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 399–402, New York, NY, USA, November 2005. ACM.
- [36] P. Stark. High resolution computed tomography of the lungs. *UpToDate*, September, 2007.
- [37] G. T. Toussaint. The use of context in pattern recognition. *Pattern Recognition*, 10:189–204, 1978.
- [38] Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, C.-T. Zhang, K. Suzuki, and K. Doi. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Medical Physics*, 30(9):2440–2454, September 2003.
- [39] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan. Computer recognition of regional lung disease patterns. *American Journal of Respiratory and Critical Care Medicine*, 160(2):648–654, August 1999.
- [40] W. R. Webb, N. L. Müller, and D. P. Naidich, editors. *High-Resolution CT of the Lung*. Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2001.
- [41] T. Westerveld. Image retrieval: Content versus context. In *Recherche d'Informations Assistée par Ordinateur (RIAO'2000) Computer-Assisted Information Retrieval*, volume 1, pages 276–284, Paris, France, April 2000.
- [42] T.-F. Wu, C.-J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.*, 5:975–1005, 2004.
- [43] Y. Wu, E. Y. Chang, K. C.-C. Chang, and J. R. Smith. Optimal multimodal fusion for multimedia data analysis. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 572–579, New York, NY, USA, October 2004. ACM.
- [44] T. Zrimec and J. Wong. Improving computer aided disease detection using knowledge of disease appearance. *Studies in Health Technology and Informatics*, 129:1324–1328, 2007.